МЕТОДЫ И АЛГОРИТМЫ ПОСТРОЕНИЯ МАТЕМАТИЧЕСКИХ МОДЕЛЕЙ ДАННЫХ В ЗАДАЧАХ МЕНЕДЖМЕНТА КАЧЕСТВА

П.С.Серенков, кандидат технических наук, доцент, заведующий кафедрой СМИС;
В.М.Романчак, кандидат физико-математических наук,
Белорусский национальный технический университет;
Э.М.Короневич, студент группы 113519 кафедры СМИС

Менеджмент качества как специфический вид деятельности предполагает, прежде всего, комплексное решение задач планирования, обеспечения, управления и улучшения качества каждого процесса и сети процессов в целом.

В соответствии с международными стандартами ИСО семейства 9000 все параметры качества изделия должны задаваться в виде номинального значения с допуском. Это означает, что каждый параметр качества должен быть представлен в виде диапазона допустимого рассеяния своих значений, который нормируется соответствующими стандартами и оформляется должным образом. Таким образом, корректировать показатель качества можно двумя путями:

1. Воздействовать на факторы, определяющие его математическое ожидание.

2. Воздействовать на факторы, определяющие его дисперсию.

Подавляющее большинство процессов можно по степени определенности отнести к категории «черные ящики». Этот класс процессов наиболее сложен с точки зрения менеджмента качества. Все широко рекламируемые, а также малоизвестные методы, техники и подходы для решения проблем качества этих процессов можно объединить под одним названием «статистическое моделирование процессов», т. к. основаны они на принципах теории вероятности и математической статистики. «Пищу» для статистического моделирования процессов составляют «данные о качестве» – накапливаемые в количественной или качественной форме данные о параметрах качества каждого процесса сети процессов [1]

Однако известные методы решения задач моделирования процессов не всегда могут дать удовлетворительное решение поставленной задачи, предоставляя некорректные модели. Анализ некорректной модели искажает картину реального состояния системы, даже если алгоритм и математический аппарат обработки данной модели будет идеален с точки зрения математики. В данной статье обсуждается гипотеза построения модели процесса с помощью глобальной рациональной интерполяции результатов экспериментов (данных о качестве) и ее использования для решения оптимизационных задач а также автоматизации процессов менеджмента качества.

Математическую модель данных о качестве можно представить в виде

$$\varphi = f(x_1, x_2 ... x_n),$$

где φ – выходной параметр (показатель качества, результативность и т.п.),

f – функция отклика,

$x_1, x_2 \dots x_n$ – факторы.

В настоящее время наиболее распространенным способом описания зависимости между параметрами оптимизации и влияющими факторами является регрессионная модель [2], суть которой лежит в нахождении коэффициентов регрессии при выполнении условия наименьшего значения суммы квадратов разностей действительных и аппроксимированных значений. В основном используется линейная регрессия, которая описывается уравнением

$$\varphi = \sum_{i=0}^{m} b_i f_i(x_1, x_2 ... x_n),$$

где φ – выходной параметр,

f_i – функция связи,

b_i – коэффициенты регрессии,

$x_1, x_2 \dots, x_n$ – факторы.

Данная модель называется линейной, потому что коэффициенты регрессии имеют степень, равную единице, функции же связи могут быть различны. На практике наибольшее применение находит линейная регрессия (функция связи выражает линейную зависимость выходного параметра от факторов):

$$\varphi = \sum_{i=0}^{m} b_i x_i \,. \tag{1}$$

Сразу бросается в глаза главный недостаток данного метода описания функции отклика: происходит приближение к заранее выбранной функции, причем выбор данной функции зачастую не обосновывается априорной информацией. В качестве примера можно привести какой-либо процесс, имеющий параболоид в качестве действительной функции отклика

$$z = 100 - x^2 - y^2,$$

причем массив имеющихся данных о качестве симметричен относительно точки экстремума. В таком случае при аппроксимации функции к виду (1) экстремум найден не будет, хотя он имеет ярко выраженный характер.

С целью минимизации методических ошибок моделирования такого рода, на наш взгляд, необходимо построить функцию связи, проходящую через все точки массива данных о качестве. В качестве решения можно использовать интерполяционные полиномы, однако существует ряд трудностей связанных с применением общеизвестных интерполяционных полиномов:

- отсутствие устойчивости интерполяционных полиномов Лагранжа и Ньютона при увеличении количества узлов интерполяции до 20;

• сложность построения функции многих переменных для интерполяции кубическими сплайнами.

Рассмотрим методы построения функции отклика, не содержащие недостатков, описанных выше.

1. Применение аппроксимирующего полинома вида

$$R(t) = \frac{\sum_i y_i \cdot f(t - x_i)}{\sum_i f(t - x_i)},$$

где $f(x) = \frac{l^2}{\left[(x)^2 + l^2\right]}$;

x_i – значения факторов в узлах;

y_i – значения выходного параметра в узлах;

l – параметр, отвечающий за степень приближения функции отклика к узлам;

t – значение фактора.

Чем меньше данный параметр, тем ближе функция к узлам (на рис. 1 приведена аппроксимация при $l = 0{,}1$).

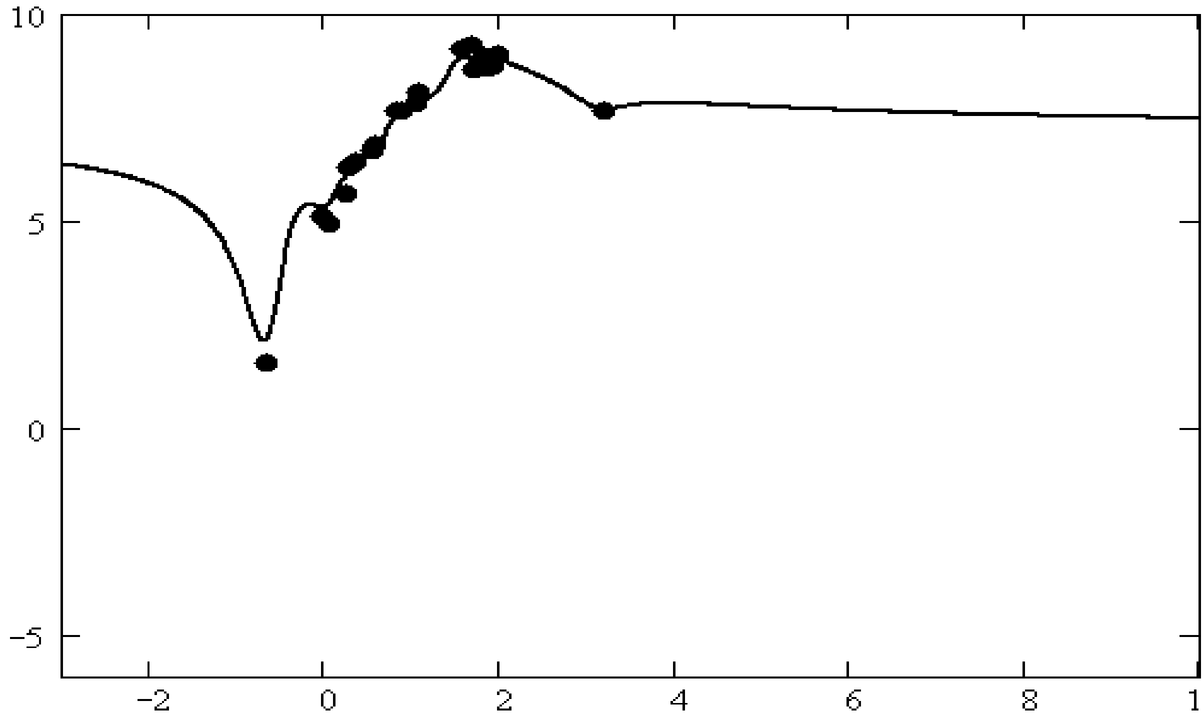

Рис. 1. Аппроксимация массива данных о качестве при помощи первого метода

Для многомерных случаев функция будет иметь вид

$$R(t_1,t_2)=\frac{\sum\limits_i z_i\cdot f(t_1-x_i,t_2-y_i)}{\sum\limits_i f(t_1-x_i,t_2-y_i)},$$

где $f(x,y)=\frac{l^2}{\left[(x)^2+y^2+l^2\right]}$,

т.е. для перехода к большему количеству переменных добавляем переменные в исходные функции.

2. Представление функции отклика в виде интегрального уравнения Фредгольма первого рода.

Интегральное уравнение Фредгольма первого рода имеет вид [3]

$$\int\limits_a^b K(x,s)\cdot z(s)\cdot ds=u(x),$$

где x – переменное значение фактора применительно к нашему случаю;

s – значения наблюдаемых факторов;

$K(x,s)$ – ядро функции;

$z(s)$ – значение некоторой заданной функции от фактора;

$u(x)$ – значение параметра оптимизации.

Возьмем ядро функции как разностную функцию $K(x-s)$, имеющую вид

$$K(x)=\frac{1}{x^2+h^2},$$

где h – параметр, определяющий степень влияния узлов в зависимости от удаленности от данной точки.

Чем больше данный параметр, тем большее количество узловых точек имеет значимое влияние на любую точку функции отклика, таким образом, чем больше данный параметр, тем более гладкой будет функция отклика.

Составляется матрица (A) значений ядра, в качестве переменных которого выступает разность между узлами факторов:

$$A = K(s_i\text{-}s_j),$$

где s_i, и s_j – узловые значения факторов (исходные данные о качестве для построения модели), *а* i и $j = 1. ..., n$.

На основании известного вектора значений параметра оптимизации можем определить вектор значений $z = \{z(s_1), z(s_2), ..., z(s_n)\}$:

$$z = A^{-1}u,$$

где $u = \{u(s_1), u(s_2), ..., u(s_n)\}$ – вектор наблюдаемых значений параметра оптимизации при факторах s_1.

Исходя из найденных значений z, составляем функцию отклика (рис. 2):

$$u(x)=\sum_i z(s_i)\cdot K(t-s_i)$$

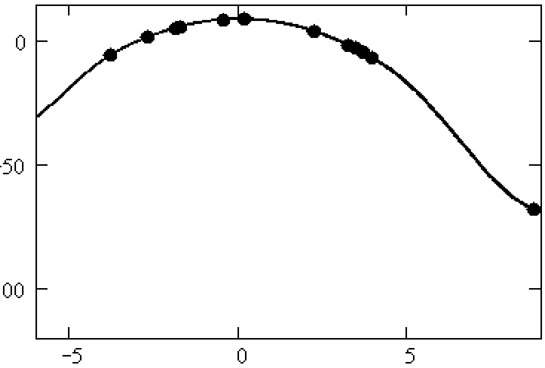

Рис. 2. Аппроксимация массива данных о качестве при помощи второго метода

Переход к многомерным случаям осуществляется путем добавления соответствующего числа переменных. Получается для ядра функции (случай для двух факторов; для большего числа факторов, все аналогично)

$$K(x, y) = \frac{1}{x^2 + y^2 + h^2}.$$

$A = K(s_i - s_j)$ преобразуется в $A = K(x_i - x_j, y_i - y_j)$,
где x_i, y_i – координаты соответствующих узловых точек s_i, а функция отклика становится функцией двух переменных (рис. 3):

$$u(x, y) = \sum_i z(s_i) \cdot K(x - x_i, y - y_i).$$

Рассмотренные примеры построения функции отклика обеспечивают построение ее математического описания по известным значениям факторов и параметра оптимизации, однако они не гарантируют точности описания рассматриваемого процесса из-за того, что сами данные о качестве могут не обладать достаточной степенью глобальности, чтобы накрыть все возможные состояния системы.

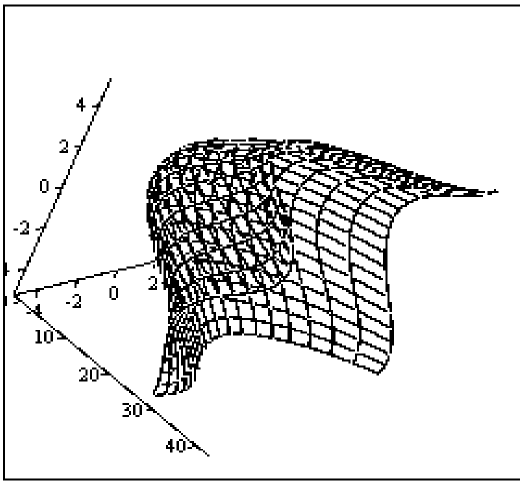

Рис. 3. Аппроксимация функции двух переменных вторым методом

Рекомендации по применению

На основании изучения рассмотренных методов можно привести ряд рекомендаций по их применению.

Аппроксимация полиномом первого вида является наиболее простым способом построения функции по известным точкам, причем функция будет приближаться к данным точкам настолько близко, насколько необходимо разработчику модели. Основным преимуществом данного полинома является то, что при его помощи можно построить функцию отклика по большому количеству наблюдаемых значений, причем чем больше рассматриваемых значений, тем лучше ведет себя функция отклика, так как полином стремится занять среднее положение в случае, когда параметр качества имеет несколько значений при одном и том же сочетании факторов, т.е. он “выбирает” статистически наиболее значимую точку. Но при небольшом количестве точек функция отклика имеет множество локальных экстремумов, что может привести к неправильному анализу данных. Поэтому данный полином лучше всего применять для построения моде-ли по накапливающимся в процессе производства данным, число которых может быть велико (тысячи).

Полином второго вида в отличие от первого обладает свойством сохранять тенденции, т.е. гладким образом продолжаться. Это хорошо при условии, что точки расположены на достаточном расстоянии друг от друга, и такая характеристика, как конечная разность (характеризует скорость нарастания функции, т.е. аналог производной), не имеет резких перепадов от больших положительных значений к большим отрицательным. Однако при резких перепадах функция начинает осциллировать. При наличии двух и более точек при одном сочетании уровней факторов матрица А становится вырожденной. Таким образом, данный вид полинома наиболее выгодно использовать при построении математической модели в планировании промышленного эксперимента, когда количество точек небольшое, причем точки обладают достаточной статистической значимостью. Применение таких полиномов для построения модели по накапливающимся в процессе производства данным является невозможным из-за показанных недостатков. Однако можно привести алгоритм первоначальной обработки массива данных о качестве для получения из набора большого числа незначимых данных меньший массив зна-чимых данных, в котором каждой области присваивается значение средней точки.

Критерием данной замены является то, что одна точка – это не есть «чистое значение» на сочетание факторов, так как в отклик входят еще и результаты влияния случайных неуправляемых факторов, и погрешность измерения. Поэтому есть смысл перейти к набору данных о качестве, обладающих какой-либо значимостью. Для получения этих данных необходимо разделить массив первоначальных данных на подобласти, а затем определить средние значения факторов и отклика в каждой из этих подобластей. Разделение массива можно произвести при помощи аппарата регрессионного анализа: выбирается минимальное количество точек (в подобласти примем его равным четырем), затем находятся коэффициенты регрессии методом наименьших квадратов для функции. Здесь мы рассматриваем случай однофакторного эксперимента

$$\varphi = b_0 + b_1 x_1 + b_2 x_1^2 ,$$

где x_1 – значение фактора;

b_0, b_1, b_2 – коэффициенты регрессии.

Если коэффициент регрессии b_2 незначим по критериям, применяемым в полнофакторном эксперименте (это означает, что набор данных аппроксимируется прямой линией), то добавляем еще одну точку в подобласть и так делаем до тех пор, пока коэффициент b_2 не становится значимым. После этого вводим новую точку, характеризующую данную подобласть, координаты которой равны средним значениям факторов и отклика в этой подобласти, далее берем следующую подобласть и проводим ту же процедуру.

Таким образом при помощи представленных алгоритмов и полиномов можно аппроксимировать любой набор данных о качестве и получить функцию отклика, характеризующую исследуемую систему.

Литература

1. Соломахо В.Л., Серенков П.С., Краснопрошин В.В. Модель «сквозного» менеджмента качества // Новости. Стандартизация и сертификация. – 2003. – № 5. – С. 65 – 69.

2. Бендат Дж., Пирсон А. Прикладной анализ случайных данных. – М.: Мир, 1969.

3. Тихонов А.Н., Арсенин В.Я. Методы решения некорректных задач. – М.: Наука, 1986.